

Publisher's version

Authors: Heimonen Tomi
Name of article: Mobile Findex: Facilitating Information Access in Mobile Web
Search with Automatic Result Clustering
Year of publication: 2008
Name of journal: Advances in Human-Computer Interaction
Volume: 2008
Number of issue: 680640
Pages: 1-14
ISSN: 1687-5907
Discipline: Natural sciences / Computer and information sciences
Language: en
School/Other Unit: School of Information Sciences

URL: <http://www.hindawi.com/journals/ahci/2008/680640/>

URN: <http://urn.fi/urn:nbn:uta-3-726>

DOI: <http://dx.doi.org/10.1155/2008/680640>

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.

Research Article

Mobile Findex: Facilitating Information Access in Mobile Web Search with Automatic Result Clustering

Tomi Heimonen

Tampere Unit for Computer-Human Interaction, Department of Computer Sciences, University of Tampere, Kanslerinrinne 1, Pinni B 2011, 33014 Tampere, Finland

Correspondence should be addressed to Tomi Heimonen, tomi.heimonen@cs.uta.fi

Received 1 April 2008; Accepted 14 September 2008

Recommended by Matt Jones

Designing an effective mobile search user interface is challenging, as interacting with the results is often complicated by the lack of available screen space and limited interaction methods. We present Mobile Findex, a mobile search user interface that uses automatically computed result clusters to provide the user with an overview of the result set. In addition, it utilizes a focus-plus-context result list presentation combined with an intuitive browsing method to aid the user in the evaluation of results. A user study with 16 participants was carried out to evaluate Mobile Findex. Subjective evaluations show that Mobile Findex was clearly preferred by the participants over the traditional ranked result list in terms of ease of finding relevant results, suitability to tasks, and perceived efficiency. While the use of categories resulted in a lower rate of nonrelevant result selections and better precision in some tasks, an overall significant difference in search performance was not observed.

Copyright © 2008 Tomi Heimonen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Mobile devices, such as personal digital assistants and mobile phones, are increasingly used for browsing mobile Internet services. This growth is enabled by the development of mobile data transfer technologies, as well as improvements in mobile World Wide Web browsers. It is estimated that the use of mobile Internet services will triple by 2013 [1]. To fuel the growth of service adoption, yearly sales of mobile devices are expected to exceed one billion in the near future [2], making them an attractive medium for various Web service providers. A recent survey reported that nearly 80% of respondents in the United States and Europe have access to mobile Web and 32% make use of mobile Web services [3]. It is apparent that this increase in the use of mobile devices and applications will change how people look for and interact with information. Mobile information services and mobile Web access will undoubtedly become as indispensable methods of information access as the Web currently is on desktop computers. A key challenge in enabling this growth is in the design of usable services—only a third of mobile Web users report being satisfied with their experience of mobile Web use [3].

The evolution of Web use on mobile devices is following a similar trend as on the desktop. Information portals maintained by mobile service operators are making way to search services that directly link to Web pages of interest [4]. While mobile search services provide information access on the go, it is the devices that pose a number of serious constraints for the design and development of services, such as their relatively small screen space, limitations posed by proprietary software architectures, and limited data transfer capabilities. It is, therefore, unsurprising that while major search engine providers and handset manufacturers have launched mobile search products of their own [5–8], the user experience of such services remains compromised when compared to the desktop. Although these services and products are designed for mobile devices, ultimately the search results themselves are in many cases presented and interacted with much in the same way as on the desktop. The search engine result pages continue to use a flat, ranked result list to present the results. Finding relevant information from these long lists can be a difficult for users who typically enter roughly two search terms per query [9] and expect the search engine to provide relevant results within the first few results. Problems inherent to the traditional search result

presentation format, such as the need for vertical scrolling, are aggravated by the interaction limitations posed by the mobile devices. Ranked result lists also fail to provide an effective overview of the themes present in the result set, forcing the users to browse page by page through the results to gain one. Results fulfilling the users' information need may remain unseen simply due to an ambiguous query that does not produce relevant results in the first few result pages—the maximum number that users typically bother to browse through [10].

Our research aims at developing innovative solutions that improve the user experience of mobile Web information search. We focus on search result evaluation, especially on how to best support users in forming an overview of the results and interacting with the entire result set, including the evaluation of individual results. This article discusses the development and evaluation of a new mobile Web search user interface concept called Mobile Findex [11]. It utilizes automatically clustered result categories for organizing and exploring search results. Designed primarily for efficient, one-handed use on mobile phones, Mobile Findex aids users in the search result evaluation process by providing access to the results using a set of representative categories, which are composed of frequently occurring words and terms in the search result summaries. Categories are used to quickly drill down into smaller, focused result sets likely to be of interest to the user. Moreover, categories also present an overview of the prevalent topics within the results, and thereby the category list can be used for evaluating the success of the whole query before committing to viewing individual search results.

We carried out a user study with the Mobile Findex prototype to investigate how the proposed concept compared to the ranked result list presentation paradigm. First, we were interested in establishing whether automatic result categories can be used to present Web search results in a mobile search user interface in a way that makes it efficient for users to identify relevant results and thus facilitate information seeking. Toward that end, we benchmarked Mobile Findex against a ranked result list interface using standard metrics such as precision and recall. In contrast to several previous studies, an actual mobile device was used in the experiment to increase the validity of the test setting. Second, we wanted to study the differences in the perceived user experience between the category-driven and ranked result list approaches. This was done by systematically collecting subjective feedback from participants during the study.

In the following, we present the design principles behind the Mobile Findex interface, followed by details of the user study and its results. We conclude by discussing the results and their implications for designing mobile Web search user interfaces and the role that categories can play. Moreover, we also highlight the need for improving the evaluation methodology so that it can better capture the subjective, experiential aspects of using mobile Web search engines for information access. This study is a limited, initial exploration of the benefits of category-based user interfaces for mobile search. Together with other studies targeting mobile Web

search experience, it can help highlight future avenues of research in the area.

2. Related Studies

Our review of previous research covers studies on both desktop Web search as well as mobile Web search as, unsurprisingly, many of the techniques used in mobile Web search interfaces can be tracked to developments in desktop search. Studies on mobile search interfaces provided one basis for our own research. It is also grounded on work done on various search result categorization approaches, some of which has also taken place in the area of mobile Web search. We will also review studies on search result presentation as they pertain to the design issues we encountered during the development of Mobile Findex.

2.1. From Desktop Web Search to Mobile Web Search

Research on Web search interfaces is a longstanding effort in the information retrieval community, and subsequently the human-computer, and more recently the human-information interaction communities. The seminal work by Jansen et al. [9] and later research by Jansen and Spink [10] provide us with a realistic view of how real users utilize Web search engines in their own information seeking tasks. They tend to use short queries, with single-term queries constituting 20–35% of queries (depending on search engine) and view only the first few pages of results (with 60–83% of users viewing only the first result page). These kinds of interactions lead to few results being considered and result in problems in finding the desired information. This in turn can lead to laborious query reformulation if the first few pages fail to produce relevant results, or outright abandonment of the search task. Aula et al. [12] have shown that problems with query formulation and operator usage are not limited to novice users, as also expert users of Web search engines struggle with their queries and result evaluation. One of their key findings is that category-based presentation of search results provides benefits to experienced users, and it could partially help overcome the problems caused by ambiguous queries. It is easy to appreciate the appeal of result categorization because of the basic human need to organize information to make it easier to process, and category-based interfaces have been proposed as one approach to providing result overviews (e.g., [13, pages 268–276]). Categories are by no means the only solution, and currently major search engines such as Google and Yahoo! provide assistance, for example, in the form of progressive query completion and alternate query suggestions as ways of improving the quality of queries, and subsequently of the search results.

Studies on mobile Web search are less numerous, although recently some research has emerged on the topic. Kamvar and Baluja [14] presented the first large-scale study of wireless search behavior. Their results, based on data gathered from a major US operator's traffic logs, mirror those reported by Jansen et al. [9], indicating some similarities in

search behavior between desktop and mobile users. Single-term queries accounted for roughly 36% of queries and the vocabulary size of queries was quite limited when compared to desktop search. Moreover, the exploration of results was much more limited than on the desktop, with only 8.5% of sessions proceeding beyond the first result page. This is understandable given the relatively higher cost of interactions in the mobile environment (e.g., difficulty of interacting with links and the associated data transfer costs). More recently, Church et al. [4] carried out a similar search log study in which they found remarkably similar results. In their data set of European mobile Internet use, 58% of queries contained two search terms or less and there was a high degree of overlap between queries. Their findings also provide interesting insights into the mobile information access behavior that users engage in. According to their results, searching constitutes only 6% of overall interactions in the mobile Web. However, users that do engage in mobile search are more active users of mobile Internet services than those limiting themselves to browsing. It is interesting to consider the explaining factors for this. Church et al. propose that it is the early adopters of mobile technology that primarily use mobile search services, and we can conjecture that these users would also be more comfortable with using search interfaces on mobile devices. Thus searching complements browsing activities as a method of information access, similarly to the early phases of Web search adoption in the early 90s.

The main reasons for the lack of search service adoption in the mobile Web would appear to be twofold: on one hand, the current data transfer pricing plans are relatively expensive, considering the availability of content suitable for and directed at mobile users. On the other hand, search engine user interfaces themselves are in need of improvement, as they need to better account for the mobile context of use. For example, the difficulties users have with mobile text entry is a known problem in general and for search especially, as noted by both Kamvar and Baluja [14] and Church et al. [4]. Moreover, as Kamvar and Baluja conclude, the perceived cost of undirected exploration appears to be too high, prohibiting users from going past the first result page should it not yield clearly relevant results. However, given the parallels in the evolution of search behavior in the desktop and mobile environments, we believe that the breadth and depth of queries will increase in the future as data transfer costs decrease and the use of mobile search becomes more widespread, aided by the development of innovative interface solutions. We believe that integrating search result categorization in the mobile search interfaces could be one key solution in ameliorating the above problems, specifically the lack of results exploration.

2.2. Search Result Categorization

Organizing search results into meaningful groups, categories of interrelated results, can help information seekers make sense of search results and decide which actions to pursue [15]. Approaches to organizing results into categories vary;

for example, we can use structural information of the document collection, document classification, or document clustering techniques to form the categories [15, 16]. Techniques that utilize structural information organize results based on the metadata associated with each document, for example, bibliographic or taxonomic classifications, or location of the document in a directory structure. Classification techniques divide documents into predefined categories based on their content, either manually or using a variety of automated methods, such as support vector machines or Bayesian classifiers. Document classification typically produces descriptive category names and meaningful conceptual hierarchies, but the classification algorithms themselves can be quite complex and end users can have problems understanding their functional principles. One of the biggest drawbacks in using classification techniques in Web search interfaces is the difficulty of creating and maintaining the classification structures and their contents in such a dynamic environment as the World Wide Web. In contrast to classification, clustering techniques form clusters of documents based on shared properties, which are derived from the textual features of the documents such as frequently occurring words or phrases. Clustering techniques can be easily automated and are applicable even for short documents or excerpts, such as search result captions (also called snippets). Since clustering is based on words and phrases from the result documents, cluster hierarchies can reveal dominant themes in the result set. Clusters can also help in highlighting likely results of interest, for example, by pointing out documents written in a foreign language [15]. One of the main problems associated with clustering techniques is labeling. Whereas classification techniques rely on category names given by humans, clustering techniques use the most frequent or distinctive words found in the documents as labels. This can result in long and incomprehensible labels that do not necessarily correspond to the content of the clusters.

The above approaches have been used to enhance result presentation in information retrieval systems and Web search interfaces. Flamenco, a hierarchical faceted metadata interface by Yee et al. [17] and automatic classification approaches, such as SWISH [16] by Chen and Dumais, provide hierarchical category structures with descriptive category labels to support the exploration of search results. In contrast, many proposed clustering approaches [18–20] produce a flat list of cluster labels. However, also hierarchical clustering techniques have been proposed, for example, Ferragina and Gulli [21] introduced SnakeT, which uses gapped sentences from text instead of single terms or phrases as labels for the result clusters. Currently, a number of commercial Web search engines utilize result clustering in their user interfaces. Implementing online clustering can be quite challenging technologically [21], which has likely prevented its widespread commercial adoption so far.

2.3. Categories in Mobile Web Search

Of special interest to our research is how well category overviews are applicable to mobile search interfaces. Chan et al. [22] proposed a system for browsing document

collections based on clustering and hierarchical document summarization. In their system, hierarchically presented concepts are accompanied with relevant sentences from the result documents to show the context in which the concept occurred. More recently, Carpineto et al. [23] introduced Credino, a clustering search engine for mobile devices based on concept lattices, a form of hierarchical clustering. In their approach, the categories are arranged as an expanding hierarchy, where the cluster labels act as links to result pages. Their user study demonstrates that search result clustering is both feasible and effective as an interaction paradigm on mobile devices, and it also provides higher performance than ranked result lists. However, their evaluation is quite limited in scope, so it is difficult to assess how well their results can be generalized to other category interfaces. Moreover, their interface design targets handheld devices with stylus-based pointing interactions. It is unclear how usable such a hierarchical clustering structure would be on a mobile phone without a touch screen, arguably the most common platform currently in use.

Coupling categories more tightly with the result list has also been considered, as it can help the user retain sense of the overall category structure while scanning the results. Buchanan et al. [24] proposed LibTwig, a category-based overview interface for mobile digital libraries. The LibTwig user interface organizes results as an expanding outline tree, which the user can explore by selecting tree nodes until the actual result documents are reached. Evaluations of LibTwig, although only indicative, suggest that nonexpert Web users prefer the outline approach because it provides them with a good overview of the result set. As with Credino, the LibTwig interface relies on stylus-based interaction and hence it might prove unwieldy when used on a device that only features a traditional device keypad and push buttons for input.

Karlson et al. [25] leveraged the keypad-based interaction paradigm prevalent in mobile phones in FaThumb, a search interface based on a hierarchical faceted metadata approach similar to Flamenco [17]. FaThumb presents result categories as a grid element, whereby each category is mapped to a button in the mobile phone keypad. This category-to-button mapping is intended to reinforce spatial and motor memory support for interactions. The design was validated in a user study, where FaThumb was found to be more suitable than keyword entry searching for exploring large, multifaceted data sets. However, it is likely that the spatial and motor learning effects can only be effectively leveraged in domains that feature relatively static category hierarchies. This limits the applicability of the FaThumb concept for Web search applications, where the category structure would have to be adapted to the contents of the query.

2.4. Summary

Search user interfaces that provide users with category-based views have been shown to offer advantages over ranked result lists. It can be concluded that the main reasons for these advantages are twofold. First, categories provide an effective overview of the whole result set, thereby giving the users a “feel” of the quality of the results. Second, categories facilitate

navigation as an interface mechanism by allowing the users to drill down into successively smaller result sets of interest. However, most current mobile phones rely on scrolling and selection using the keypad and multiway navigation key as the main interaction methods, which limits the interaction design space of category-driven search interfaces. Many of the search interfaces discussed above base their interaction model on direct manipulation via a stylus—which makes it difficult to apply their prominent features in scenarios where one-handed use is necessary or desirable, either due to device in question or the context of use. In the interaction design of Mobile Findex, we wanted to take advantage of categories to provide effective overviews, while also addressing the needs of one-handed use. This led us to adopt progressive disclosure as a guiding principle in the design, which will be explained further in the following sections.

3. Mobile Findex

In order to be able to evaluate the proposed category-based mobile Web search interface concept in user tests, we developed a custom software experimentation platform. The resulting Mobile Findex mobile search application framework consists of two main components: server-side search result clustering engine and mobile client application. The clustering engine and client application communicate over an HTTP connection using a custom protocol. In the following, we briefly describe the underlying Findex clustering engine and continue with an indepth description of the proposed search user interface concept and its design rationale.

3.1. Findex Search Result Clustering Algorithm

We use the Findex clustering algorithm [19] and its software implementation, the clustering engine, to execute search queries and generate result categories. The clustering engine is implemented as a Java component that can be integrated into both standalone applications and Web services. The engine executes search queries, processes the results into categories, and sends them to the client application. It is also possible to use cached results, for example, in experiments that require a static dataset across queries and participants. The communication and clustering components of the engine are functionally separated from the search engine component; it is possible to use any search engine as the underlying data source, provided that it features a suitable application-programming interface (API). The current Mobile Findex implementation uses the simple object access protocol (SOAP) version of Google Web API.

The particulars of the clustering algorithm are described in more detail elsewhere [26, pages 42–46]. The algorithm employs a fairly straightforward document clustering technique: it uses word and phrase frequencies in the search result captions (snippets) as the dominant factor in forming a set of categories. Because the algorithm and resulting cluster labels are based on word and phrase occurrences in the text, it is fairly easy for nonexpert users to understand the functioning of the algorithm. We believe that by understanding how the

underlying clustering mechanism works, the users are better able to utilize the categories it produces. This understanding may alleviate some of the concerns raised by Hearst [15] on the mismatch between cluster labels and the contents of the results within the clusters.

The clustering algorithm has three main stages: (1) text trimming, (2) category candidate extraction, and (3) redundancy-filtering. In the first stage, stopwords and other nonalphanumeric strings are removed from the results. Next, the algorithm extracts potential candidates from the snippet text by using a “moving window” approach, thus effectively compiling a list of all possible consecutive words and phrases present in the text. In the redundancy-filtering step, the algorithm iteratively removes category candidates that are composed of the same words (e.g., “Stanford University” and “University Stanford”) and phrases that are subphrases of longer candidate phrases. In the end, the most frequently appearing candidate phrases are selected. The content of the results in each category contain one or more occurrences of the category phrase. The categories are not mutually exclusive and, therefore, some results may appear in multiple categories.

There are certain limitations to the clustering algorithm. The quality of the clusters is obviously dependent on the content of the search result captions. Since no query-biased processing is applied while extracting the category candidates, the algorithm can also result in “out of context” labels, which do not seem to directly relate to the query in anyway. Another common problem is excessively broad, generic labels. The former are categories that seem relevant but do not convey any information about their context (e.g., “elections” for the query “Iraq”), and the latter in categories that are too general to convey any meaning (e.g., “world”).

3.2. Mobile Cluster-Based Search Interface Design

Several guidelines exist for designing interactions in mobile user interfaces. We used the seminal guidelines for designing mobile search user interfaces proposed by Jones et al. [27] and Jones and Marsden [28] as a starting point for the design process of the Mobile Findex user interface. As such, the two main goals of our design were to allow users quickly evaluate the success of their queries and subsequently give them enough information about individual results to make judgments on their usefulness. Jones and Marsden suggest the use of overviews, either in the form of automatic clustering categories or predefined, topical categories, as a solution for the first design goal. Accordingly, Mobile Findex uses automatically generated clusters both to provide an overview of the result set and to act as filters that narrow down the amount of results shown. The cluster labels and the corresponding search results are split into separate views, both to minimize the need for vertical scrolling and to maximize the use of the limited display space for presenting information about the search results at each stage. This solution is reminiscent of the approach proposed by De Luca and Nürnberg [29], in which the search results are

presented in abbreviated form in an initial result view and in full, annotated form in a detailed results view. Their interface relies on stylus input, so the concept was not directly applicable for our design. We also considered integrating the category list in the initial search screen alongside the query box in order to streamline the interaction. In the end, we decided against it, as we could not come up with a satisfactory and efficient solution for focus switching between the query field and the category list, something that would be a trivial challenge if designing for touch screen devices.

Mobile Findex presents results in a dynamically expanding result list (in the vein of WaveLens [30] by Paek et al.). The goal is to provide the users as much information about the search results as possible in the limited space available while attempting to further reduce the amount of scrolling. Results are presented as a combination of the original unmodified title, result caption, and URL for the item currently in focus. Results above and below the focused item only display the title and the URL. We also considered the possibility of dynamically altering the content of the results, for example, by using representative key phrases instead of text captions [31], using caption texts of varying lengths [32], using different text processing schemes when constructing the content of the captions [33], or visualizing the occurrences of the query terms in the result document (e.g., [34, 35]). These alternatives were ultimately discarded during the design process to avoid overloading the interface with new features in this initial stage of exploring the design space. Further studies on how to effectively display categories and the metadata related to individual results are needed to further map the design space.

The resulting Mobile Findex user interface (Figure 1) consists of three distinct views: the query view, the category view, and the result list view. Navigation in the interface takes place by using the multiway navigation key or arrow keys common to most modern mobile phones. The top element in each view changes to highlight the currently active view, and it also provides contextual information, such as the search query or the selected category name. The query view resembles a typical search user interface, containing an input field for entering query terms. The category view is used to present the categories. Each row in the category list consists of the label (which can span multiple lines) and a numerical indicator showing the number of results contained in that category. An additional item titled “all results” is included at the end of the list, and it can be used to access all results for the query and thereby bypass the categories altogether. The result view presents the individual results, in the ranking order of the underlying search engine, using the focus-plus-context visualization discussed previously. The focused item displays the title and the URL in their entirety and up to three lines of the caption. Items in the context area only display a shortened title and URL. This abbreviated format allows users to review more results at a time, especially in the initial view, while deciding whether to investigate the selected category further.

Design guidelines [27, 28] also stress the importance of effective interaction. Toward this end, Mobile Findex

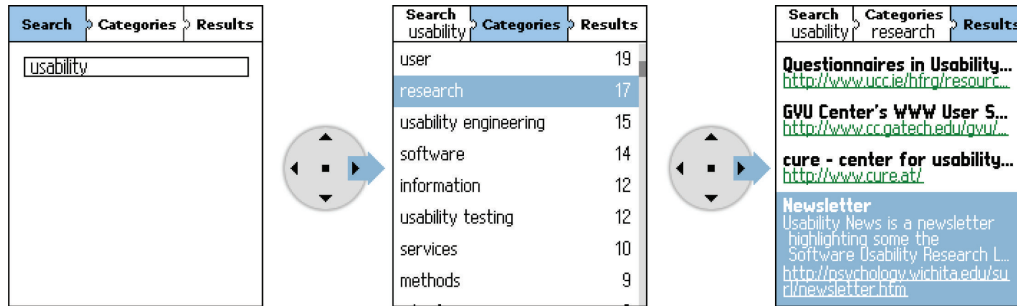


FIGURE 1: Mobile Findex search user interface, with the query view (left), category view (middle), and result list view (right).

provides a streamlined interaction model, whereby all the functions of the user interface can be accessed using the multiway navigation keys. Navigation between the views takes place with left-right selections and scrolling through the lists with up-down selections. Individual results can be selected by pressing down on the navigation key, after which the phone's built-in web browser application is launched to present the resulting web page. This design choice was made out of necessity, as the Java MIDP platform currently lacks a suitable user interface component capable of displaying HTML content.

4. User Study

We evaluated the Mobile Findex concept in a mobile Web search scenario. The main goal of the evaluation was to compare Mobile Findex to a mobile Web search interface using a ranked result list. The evaluation was organized as a laboratory experiment, in which the different factors affecting the usage situation could be controlled. The experimental setting is based on a previous experiment by Käki and Aula [19], in which a category interface was compared to ranked result lists in the context of desktop Web search.

4.1. Participants

A total of 16 (8 female, 8 male) participants volunteered for the study. They were all undergraduate students at a local university aged between 21 to 33 years ($M = 23$). All participants had considerable experience in using computers (7–19 years, $M = 12$), mobile phones (3–11 years; $M = 8$), the Web (4–10 years; $M = 7$), and Web search engines (4–10 years; $M = 6$). All used computers and the Web daily. Web search engines were used daily by 8 and many times a week by 7 participants. The web search engine of choice was Google (15 out of 16 participants), with one participant reporting the use of the built-in, user-selectable search engine functionality of the Mozilla Firefox browser. None of the participants had any significant experience in using mobile Web search engines. While ideally we would have liked to include participants with mobile Web search experience, it proved extremely difficult to find people with such experience at the time of the study. However, the participants do, otherwise, fit the early adopter profile given their overall technological expertise.

4.2. Method

The experiment was organized as a within-subjects design with one independent variable *user interface* with two levels: Reference UI (the ranked result list user interface) and Mobile Findex UI (the Mobile Findex category-based user interface). The following dependent variables were measured in order to evaluate the performance of the user interfaces: (1) task duration in seconds, (2) number of result selections per task, and (3) relevance of selections per task. The participants' subjective views toward the user interfaces were elicited using two questionnaires, administered after they had completed tasks with each interface. A final questionnaire comparing the interfaces was administered at the end of the experiment.

During the experiment, the participants were asked to carry out a total of 12 information-seeking tasks, divided into two thematically balanced blocks of 6 tasks each. One block of tasks was carried out using the Reference UI and the other using the Mobile Findex UI, resulting in four distinct UI—task block combinations. The order in which the combinations were presented was counterbalanced between participants to eliminate learning effects. The order of tasks within blocks was randomized. This resulted in a total of 192 ($16 \times 2 \times 6$) task level observations being recorded during the experiment.

4.3. Tasks

The tasks used in the experiment were information-seeking tasks, with the overall goal of finding results pointing to Web pages that fulfill a specific information need [36]. The task topics covered a variety of themes, for example, general interest, shopping and historical events. The task descriptions and matching queries presented to the participant are listed in Table 1. The tasks were predominately drawn from a pool of tasks used in our previous Web search experiments.

We used the top 150 search results for each query provided by Google. The results were cached on the server to avoid introducing any changes in the result sets during the experiment. Although using predefined queries and cached results lowers the fidelity of the setting, it enabled us to draw comparisons between the interfaces. This approach is also used in previous studies comparing search user interface designs (e.g., [16, 19]). No special query operators, such as

TABLE 1: Queries and task descriptions (queries marked with asterisk are translated from Finnish).

Task description	Query
Find information about colored contact lenses.	Contact lenses
Imagine that you are buying a DVD player. Find pages with information about prices.	DVD player*
Find pages with information regarding the prevention of influenza.	Influenza
Find images of the planet Jupiter.	Jupiter
Imagine that you are buying a camera phone. Find pages with information about prices.	Camera phone*
Find pages about the University of Oulu.	Oulu
Find pages about the city of Oulu.	Oulu [used as the query in two tasks]
You suspect that you have sighted a goldeneye. Find pages with information that can help you confirm the sighting.	Goldeneye*
Find information about the Titanic shipwreck.	Titanic
You suspect that you have sighted a barnacle goose. Find pages with information that can help you confirm the sighting.	Barnacle goose*
Find images of the planet Venus.	Venus
Find information about the terrorist attack on the World Trade Center.	World Trade Center

Boolean logic or parentheses, were used as a part of the queries, given their very low popularity in general use (only about 3% of queries contain special operators [4]). The results were organized beforehand into 15 categories for each query using the Findex clustering.

During the experiment, task descriptions and queries were presented to the participants with a 1024×768 pixel display resolution, full-screen desktop application running on a Pentium 4 level Windows XP workstation. The application user interface included controls required to advance through the experiment without any moderator involvement. When user input was required, the participants controlled the desktop application with a mouse.

4.4. Reference Mobile Web Search User Interface

The implementation of the benchmark reference user interface resembles Google mobile search [6] in terms of content and functionality (Figure 2), as it appeared at the time of the study. Each result is presented as a combination of title, caption, and URL address. In addition, a number denoting position in the ranked result list precedes each result.

With this user interface, focus selection in the list is moved with up-down presses of the multiway navigation key. Movement between result pages is carried out using the “previous” and “next” links, situated at the bottom of each page. The search results are distributed across 15 result pages, with 10 results per page. Additionally, the top of the view on each page contains the query and range of displayed results.



FIGURE 2: Reference UI showing the first page of results.

4.5. Apparatus

Participants carried out the tasks with a Nokia 6680 mobile phone [37]. It features a high-color display with 176×208 pixels screen resolution and 3rd generation (3G) mobile data transfer capability. Text entry on the phone is handled with a standard nine-key keypad. Both the Reference UI and the Mobile Findex UI applications were implemented using the Java MIDP (mobile information device profile) version 2.0 application development framework.

Three different sources of information were used to record data during the experiments. The participants' interactions with the user interfaces were logged on the mobile device and transferred to a storage server for later analysis.

Task durations were registered by the desktop application and merged with the mobile interaction log during analysis. In addition, paper questionnaires were administered to elicit the participants' subjective views on the evaluated user interfaces.

4.6. Procedure

To begin, the participants were explained that the purpose of the test was “to evaluate two mobile information search interfaces,” and given instructions on how to use both mobile applications as they completed two exercise tasks (one per interface). They were also introduced to the desktop application controlling the pacing of the experiment. Next, the procedure of the experiment was described and the participants were instructed to “mark as many relevant results as possible, as fast as possible” within the given time limit. The maximum time for completing the task was limited to three minutes in an effort to reproduce a more realistic usage scenario, where the participants would be forced to find a balance between speed and thoroughness. Informal observations in previous studies have shown that if the participants are allowed to spend as long as they wish when completing information-seeking tasks, they tend to prioritize thoroughness over speed. However, in a real situation there would be other factors, such as time constraints, the importance of the information need, and the usage situation itself, which would limit the available spent on task. Participants were encouraged to utilize their own information-seeking strategies and no acceptable minimum number of selected results was given. During the experiment, the participants were not able to open the actual Web pages pointed to by the URL address of the result. This limitation was implemented to constrain the result evaluation process to the search user interfaces and their functionality.

The test moderator executed the query to initiate the task and handed the mobile phone back to the participant when all results had been received. After receiving the phone, the participant was instructed to read the task description, push the “start” button on the desktop interface and then proceed to complete the task. Likewise, upon completing the task, the instruction was to hand over the phone to the moderator and push the “done” button on the desktop interface in order to proceed to the next task. If the time limit expired during the task, the desktop application automatically ended the task and notified the participant.

Each participant completed the tasks in two blocks of six tasks: first block with one interface and then the second block with the other interface. After each block the participant was administered a questionnaire regarding the user interface. After all tasks were completed, the participants answered a questionnaire comparing the two user interfaces, as well as a background questionnaire collecting demographic information.

The functionality to mark results was added to both user interfaces. The participants were able to tag results as relevant by clicking the multiway navigation key. The selection could be removed by clicking again on a selected result. Selected

results were distinguished from other results with a visual cross-shaped marker (Figure 3).

5. Results

In the following, we present the results from the user study. Discussion of the results is divided into three categories: speed measures, accuracy measures, and subjective measures. Speed measures reflect the efficiency of use, accuracy measures the effectiveness of use, and subjective measures the perceived user experience and satisfaction.

5.1. Speed Measures

Task completion times were calculated from the moment the participant pushed “start” button in the desktop application to the moment they pushed “done”. Average task completion time for Reference UI was 130 seconds (SD = 49) and 138 seconds (SD = 42) for Mobile Findex UI. The participants completed 53% of the tasks under the allotted three-minute time limit with Mobile Findex UI and 64% with Reference UI. Search speed was calculated as the ratio between results selections and task completion time. With Reference UI, the participants collected on average 4.2 results per minute (SD = 2.0), whereas with Mobile Findex UI the rate was 3.6 results per minute (SD = 1.3). We did not observe a statistically significant effect of user interface in either case.

5.2. Accuracy Measures

The relevance of each individual result for each task was assigned prior to the experiment on a three-step scale (relevant-related-nonrelevant). This ranking was done based on the document summaries provided by the search engine. Each task was designed to contain two facets of information need: the general area of interest (e.g., the planet Venus) and specific information need (e.g., images of the planet). A result was judged relevant if it contained information pertaining to both facets, related if it only contained information pertaining to the general area of interest and nonrelevant if neither criterion was met. These ratings were the basis for calculating three distinct accuracy measures: *precision*, *recall*, and *qualified search speed*. Precision and recall are two de facto metrics used in the evaluation of information retrieval applications such as Web search engines. Qualified search speed is a proportional measure that takes task duration into account when calculating precision and is thus a more sensitive measure for accuracy than precision [38].

5.2.1. Precision and Recall

Precision was calculated as the proportion of relevant result selections among all results selected. Average precision with Reference UI was 48% (SD = 12) and 53% (SD = 10) with Mobile Findex UI. While the overall difference in precision is not significant, on task level significant differences in precision were observed in four tasks (corresponding to queries “DVD player”, “Jupiter”, “camera phone”, and “Oulu”

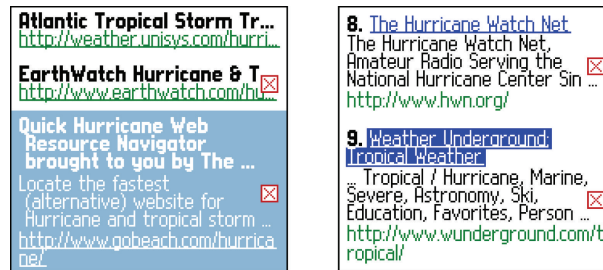


FIGURE 3: Selection markers in Mobile Findex UI (left) and Reference UI (right).

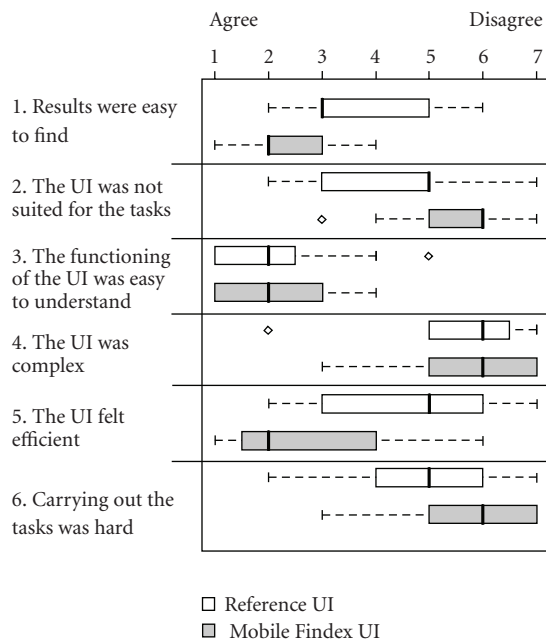


FIGURE 4: Subjective ratings of Reference UI and Mobile Findex UI.

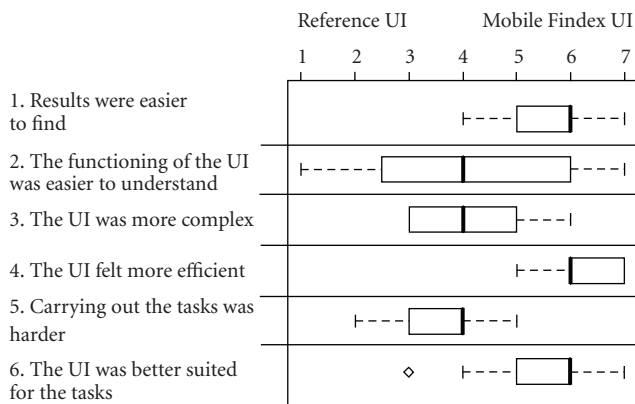


FIGURE 5: Subjective ratings comparison between the Reference UI and Mobile Findex UI.

[city]). Using Mobile Findex UI resulted in higher precision in the first three tasks, whereas Reference UI resulted in higher precision in the fourth task. Table 2 shows task-

specific precision percentages and results from independent samples *t*-test. The data from one participant was not included in the analysis of the “DVD player” task due to irrecoverable data corruption in the interaction log.

Recall was calculated as the proportion of relevant results selected by the user to all relevant results in the result set. The average recall for the participants with both the Reference UI and Mobile Findex UI was 21% (SD = 9 and SD = 8, resp.). The differences in recall between the user interfaces were not statistically significant.

5.2.2. Qualified Search Speed

Two qualified search speed measures were calculated rate of acquiring relevant results and rate of acquiring nonrelevant results. With Reference UI, the average rate of acquiring relevant results was 2.0 relevant results per minute (SD = 1.4) and 1.9 relevant results per minute (SD = 0.9) with Mobile Findex UI. User interface did not have a significant effect on the rate of acquiring relevant results. However, a comparison of nonrelevant result acquisition rates is more interesting: the participants made 1.1 nonrelevant selections per minute with Reference UI (SD = 0.8) and 0.7 with Mobile Findex UI (SD = 0.4). Significant effect for user interface was observed $t(15) = 2.32, P < .05$.

5.3. Subjective Measures

The participants were presented with subjective evaluation questionnaires during the experiment to measure their experiences. After completing tasks with one user interface, they filled in a questionnaire with six claims pertaining to it. The claims covered their views on the perceived efficiency and effectiveness of use. Each claim was answered using a seven-point scale that ranged from *agree* (1) to *disagree* (7). Figure 4 presents the answers for each claim as box-and-whiskers plots, showing the interquartile range, extent of values (1.5 times the IQR) and median.

Overall, the participants’ subjective ratings of the two user interfaces differed significantly on three claims and in all cases the difference was in favor of Mobile Findex UI: (1) results were easy to find, (2) the UI was not suited for the tasks, and (5) the UI felt efficient. Analysis of the answers using exact Wilcoxon signed-rank test gives $V = 105, P < .01$; $V = 15.5, P < .05$; and $V = 106, P < .01$, respectively.

TABLE 2: Task-specific precision and significance.

Query	Reference UI	Mobile Findex UI	Significance
DVD player	23%	54%	$t(13) = -2.9, P < .05$
Jupiter	43%	74%	$t(14) = -2.5, P < .05$
Camera phone	15%	36%	$t(14) = -2.6, P < .05$
Oulu [city]	91%	67%	$t(14) = -2.7, P < .05$

At the end of the experiment, the participants filled in a questionnaire that contrasted the two user interfaces, using the same claims as above (presented in a different order). Each claim was answered using a seven-point scale that ranged from *Reference UI* (1) to *Mobile Findex UI* (7). Figure 5 presents the answers for each claim as box-and-whiskers plots, showing the interquartile range, extent of values (1.5 times the IQR) and median.

The participants' answers differ significantly from the hypothesized median (4 = no perceived difference between user interfaces) on three claims: (1) results were easier to find, (4) the UI felt more efficient, and (6) the UI was better suited for the tasks. The differences are statistically significant and in favor of Mobile Findex UI. Mann-Whitney U test gives $U = 248, P < .001$; $U = 256, P < .001$; and $U = 224, P < .001$, respectively.

6. Discussion

This study attempted to answer two research questions focusing on support mechanisms for mobile Web information access. The first was whether automatic result categories could be integrated into a mobile Web search user interface in a way that facilitates efficient information seeking. The second research question was to find out how the proposed Mobile Findex user interface compares to a ranked result list search interface in terms of perceived user experience. Our evaluation of Mobile Findex in a Web search experiment conducted with an actual mobile phone and using representative Web search tasks provided answers to these questions.

6.1. Categories Improve Search Performance in Certain Situations

Results from task completion measures do not show a clear difference between the two search user interfaces. We could find subtle differences in result selection performance between the two interfaces using standard evaluation metrics, such as time to complete task and result selection speed. The participants completed search tasks on average 6% faster and their overall rate of result selection was 17% higher with the ranked result list user interface. However, no significant effect for user interface was observed in either case. One approach to explain this result is to consider the differing styles of interaction the interfaces facilitate. Mobile Findex, designed around a result-filtering paradigm that relies on back and forth navigation, may have encouraged the

participants to explore the result set in more detail than the reference interface. Conversely, in the reference user interface interaction was mostly serial, from one result page to the next, using links at the bottom of the result list. The ease of exploration that categories provide comes at the expense of time and overall task performance. We can draw certain design implications from this observation. It is likely that the context switching users must engage in when going from categories to results will limit the effectiveness of category-based interfaces from a purely performance standpoint. One solution is to integrate categories into the result list itself, by organizing the list into a visual, interactive hierarchical structure. Perhaps a more suitable approach for the scenarios considered in this article is to provide navigation aids, such as an on demand category selector in the result list, to facilitate easier switching between categories.

In terms of overall effectiveness, the participants made a higher proportion of relevant result selections with Mobile Findex (53% versus 48%), although the difference is not significant. In individual tasks, where a significant difference was observed, the explanation relates to the content of the result clusters. For example, in the task where the participants were instructed to “find images of the planet Jupiter,” cluster labels contained the entry “images,” enabling the participants to directly drill into a set of results likely to contain links to image sites. Similarly, in the tasks in which the participants were asked to find pricing information about DVD players and subsequently about camera phones, the clusters contained entries for “price,” which provides a focal point to start the exploration of results. This finding intrigued us as it reflects the kind of activities people might likely engage in with mobile Web search when, for example, window shopping and using mobile search to find pricing information, or check whether the price of a product at a store is lower than when ordered online. The case where categories failed is likewise interesting. When asked to find pages about the city of Oulu, the participants performed worse with Mobile Findex. This result we attribute to the nature of the clustering algorithm and the known tradeoffs related to cluster labeling. In this case, the clusters titled “Oulu city” and “Oulu Finland,” which sound valid considering the task, contained only two results directly relevant to the city. It is possible that seemingly relevant cluster titles may in some cases mislead the users to expect they will find relevant results within. We hope to tackle this issue in the longitudinal studies to find out whether and to what extent it negatively affects use when people use the application in their daily information seeking tasks. It is also possible that the experimental setting and preconstrued tasks

change how people approach result evaluation. Observing their usage patterns “in the wild” should provide a more complete picture of how categories are utilized.

Qualified search speed, or the rate of acquiring results, indicates that there is no practical difference in the acquisition rate of relevant results. The user interface did have a significant effect on the rate of making nonrelevant result selections. Given a three-minute search session, Mobile Findex users would make on average one nonrelevant selection less compared to the reference user interface. While the difference sounds trivial, this finding provides some evidence that similarly to desktop use [19], also in mobile use categories can be used effectively to filter out clearly nonrelevant results. This might benefit frequent searchers over a number of sessions, but a long-term study is required to observe the full effect.

Based on the performance measures, the answer to the first research question is a qualified “yes”—both user interfaces provided similar levels of performance in terms of precision and rate of result selection. The inability to show tangible performance benefit from result categorization is nevertheless surprising. Previous studies have shown that similar category-based user interfaces are superior to the ranked result list in the desktop environment [16, 19]. Clear performance improvement is also cited in a recent study of a mobile clustering search interface [23]. It appears that on the desktop the category user interface primarily draws its benefits from mouse-based interaction that enables quick swaps between categories and the ability to see categories and results in the same view. This suggests that effective use of categories in mobile interface requires the users to utilize a method of trial and error in browsing through potentially useful categories. In the current mobile search prototype, the list of cluster labels is not visible when the result view is selected. When users switch back to the category view, they must first rescan the category labels to orient themselves and find direction for the next category selection. In addition, switching between different categories requires the extra step of returning to the category view, which also increases time on task and makes it difficult to quickly compare differences in content under similar category labels. In this particular design, the benefits provided by the proposed category interface were not great enough to overcome the performance penalty incurred by the multiple views navigation.

6.2. Users Prefer Category-Based Interface Due to its Perceived Effectiveness

While the performance measures do not offer a clear picture of the differences between the user interfaces, subjective feedback provides answers to the second research question related to the perceived differences in user experience. The most apparent difference between the two interfaces is evident in the participants’ views on the efficiency of use and ease of finding results. This effect was strong both when rated individually and when the two interfaces were directly contrasted. We do not find this result particularly surprising.

Despite its tradeoffs, the proposed Mobile Findex interface provides a more convenient and engaging way to browse search results than the page-by-page navigation in the ranked result list. It is also interesting to note that the participants rated Mobile Findex higher in terms of perceived efficiency, although significant difference in performance was not measured. This suggests that the ability to get an overview of the results and being able to actively filter and narrow the result set are more essential elements of user experience than the actual level of search performance. Despite their lack of previous experience with mobile Web search, the participants rated both interfaces as relatively simple and easy to learn. This finding is supported by our informal observations during task completion. Due to experimental considerations we were not able to include query formulation and reformulation stages of search. Although categories do not actively support query formulation, category labels can suggest new query terms. A future direction to pursue would be studying whether we can support the query formulation process with the use of categories, for example, by providing a one-click option of adding the label to the current query.

The participants found the ranked result list interface to be less suited for search tasks than Mobile Findex. This is likely influenced by the nature of the tasks that were aimed to emulate likely mobile Web search scenarios—in many cases the category labels contained keywords of interest that allowed the participants to concentrate on potential result candidates, instead of having to scroll through a long flat list page by page. Again, we can see that the measured, objective performance does not necessarily correlate with the perceived experience, prompting concerns about the use of traditional information retrieval metrics in comparing search interface designs. It should be noted that this study targeted a specific type of information seeking tasks. Current mobile operator portals are focused on supporting resource-driven search and providing access to local services, where the user’s goal is to obtain some resource, such as entertainment in the form of video clips, information about current events, or the address of a local business. Although Mobile Findex can to a degree support these kinds of activities, it is primarily designed to support general information seeking from Web content.

6.3. Suitability of Current Methods for Evaluating Mobile Information Access

During the course of the study, and also in our previous investigations, we have come to note the difficulty in adapting methods steeped in traditional information retrieval methodology to studying the user experience of Web search interfaces. This sentiment is echoed also by Carpineto et al. [23], who note, “it is not easy to evaluate the retrieval performance of a hierarchical clustering engine in a precision/recall style”. More generally, it has also been found that efficiency and effectiveness have low correlations with user satisfaction [39], which raises a concern on how to best

utilize these different measures in evaluating search interfaces and interpreting the results.

Although performance metrics cannot be wholly disregarded when evaluating search interfaces, we consider methods that gauge user satisfaction and perceived outcomes of user interactions more robust in their ability to provide insight into the information access process. One approach we would like to focus on in the future is choice-based evaluation, in which the users' explicit feedback in questionnaires and implicit feedback during interaction (e.g., when given choice, which interface they use and whether this preference changes over time) provide the basis for the analysis [40].

6.4. Limitations of the Current Study and Future Work

Effective presentation of and interaction with mobile search result categories is affected by various factors. For example, the categorization algorithm and its properties, interaction possibilities afforded by the target platforms, and the content domain all pose challenges on design. It can be difficult to tease apart the performance provided by the categories themselves and how they are arranged in the interface. In our case, the categories are formed using the Findex clustering algorithm that produces a flat category list. Utilizing a different algorithm would undoubtedly change the content of the categories and thus affect performance—unfortunately experimenting with various clustering algorithms was beyond the scope of this study. Our evaluation compared a multiple view interface based on a flat category structure to the traditional, single-view flat result list. Furthermore, we chose to limit the design space to interface solutions that would yield themselves to efficient use with the phone keypad alone. It would be interesting to follow up on this study with an evaluation that compares different clustering algorithms using the Mobile Findex user interface to gauge their relative effectiveness. A natural continuation to this study would be an evaluation of alternative presentation and interaction paradigms paired with the same clustering algorithm.

Laboratory studies with limited user samples have certain inherent limitations with regards to ecological validity and the ability to generalize the results. Moreover, we constrained the experimental design to enable meaningful comparisons between the user interfaces by using predefined tasks, queries, and result sets. The procedure also limited the participants' interactions with the results to the extent that they could not view the actual resulting Web pages. A more realistic evaluation setting is needed to form an understanding of how Mobile Findex is integrated into users' own information seeking activities, in a real mobile context of use. Toward this end, we are currently planning to release a Web-based mobile search interface based on the Findex algorithm. Further work is also needed on studying the strategies and goals of mobile searches to pinpoint the kinds of search tasks that are unique to mobile Web search. While the large-scale log analyses [4, 14] can reveal overall trends at the query level (e.g., the decline in prominence of media download-related queries), they cannot adequately inform

us about the users' intent or give insight into the result evaluation process beyond click-through data.

7. Conclusions

Mobile Web search is developing through similar stages as desktop Web search was in the late 90s. There is a current need to support mobile Web search with better interface and interaction solutions, as the field as a whole is still rapidly evolving. We presented Mobile Findex; a new mobile Web search user interface featuring automatically computed result clusters. It was evaluated in a user study with 16 participants, where search performance and user experience were measured. The participants preferred the category-driven interaction of Mobile Findex to the traditional-ranked list browsing of search results. Mobile Findex was in their view more efficient, facilitated the finding of results better, and was better suited for the search tasks than ranked result lists. This can be attributed to the key design drivers of the Mobile Findex interface: the ability to provide an informative overview of the results and a flexible way for exploring the results. While the use of Mobile Findex resulted in a slightly lower rate of nonrelevant result selections and higher precision in a number of individual tasks, an overall significant effect of user interface on search performance was not observed.

This initial laboratory study focused on comparing a search interface built around automatically computed search result categories to the traditional-ranked result list. Longitudinal field studies should be conducted to observe how category-based search interfaces are used in mobile Web search activities, and learn how they could be further improved to better meet the needs of mobile information seekers.

Acknowledgments

This work received funding from the UCIT Graduate School for User-Centered Information Technology in Finland and was supported in part by the Finnish Funding Agency for Technology and Innovation (Project no. 40279/05). Previous work by Mika Käki on the Findex clustering engine was essential in enabling this research, and we owe him a great debt of gratitude. Appreciation is also extended to Kari-Jouko Räihä for his support during the study and to Juuso Kanner for his work on the initial Mobile Findex application architecture. The author would like to thank the anonymous reviewers for their comments on the manuscript.

References

- [1] Forrester Research. European Mobile Forecast: 2008 to 2013, March 2008, <http://www.forrester.com/Research/Document/Excerpt/0,7211,42199,00.html>.
- [2] Gartner. Forecast: Mobile Terminals, Worldwide, 2000–2009, July 2005, <http://www.gartner.com/DisplayDocument?id=483696>.

- [3] Online Publishers Association. Going Mobile: An International Study of Content Use and Advertising on the Mobile Web, March 2007.
- [4] K. Church, B. Smyth, P. Cotter, and K. Bradley, "Mobile information access: a study of emerging search behavior on the mobile Internet," *ACM Transactions on the Web*, vol. 1, no. 1, article 4, 2007.
- [5] AOL® Mobile Search, <http://mobile.aol.com/aolproducts/mobile-search>.
- [6] Google Mobile, <http://www.google.com/mobile>.
- [7] Live Search, <http://m.live.com/search/>.
- [8] Yahoo! oneSearch™, <http://mobile.yahoo.com/onesearch>.
- [9] B. J. Jansen, A. Spink, and T. Saracevic, "Real life, real users, and real needs: a study and analysis of user queries on the Web," *Information Processing and Management*, vol. 36, no. 2, pp. 207–227, 2000.
- [10] B. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing & Management*, vol. 42, no. 1, pp. 248–263, 2006.
- [11] T. Heimonen and M. Käki, "Mobile index: supporting mobile Web search with automatic result categories," in *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '07)*, pp. 397–404, Singapore, September 2007.
- [12] A. Aula, N. Jhaveri, and M. Käki, "Information search and re-access strategies of experienced Web users," in *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*, pp. 583–592, Chiba, Japan, May 2005.
- [13] M. A. Hearst, "User interfaces and visualization," in *Modern Information Retrieval*, R. Baeza-Yates and B. Ribeiro-Neto, Eds., pp. 257–323, ACM Press, New York, NY, USA, 1999.
- [14] M. Kamvar and S. Baluja, "A large scale study of wireless search behavior: Google mobile search," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, vol. 2, pp. 701–709, Montreal, Canada, April 2006.
- [15] M. A. Hearst, "Clustering versus faceted categories for information exploration," *Communications of the ACM*, vol. 49, no. 4, pp. 59–61, 2006.
- [16] H. Chen and S. Dumais, "Bringing order to the Web: automatically categorizing search results," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '00)*, pp. 145–152, Hague, The Netherlands, April 2000.
- [17] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted metadata for image search and browsing," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '03)*, pp. 401–408, Fort Lauderdale, Fla, USA, April 2003.
- [18] D. R. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey, "Scatter/gather: a cluster-based approach to browsing large document collections," in *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '92)*, pp. 318–329, Copenhagen, Denmark, June 1992.
- [19] M. Käki and A. Aula, "Findex: improving search result use through automatic filtering categories," *Interacting with Computers*, vol. 17, no. 2, pp. 187–206, 2005.
- [20] O. Zamir and O. Etzioni, "Grouper: a dynamic clustering interface to Web search results," in *Proceedings of the 8th International World Wide Web Conference (WWW '99)*, pp. 1361–1374, Toronto, Canada, May 1999.
- [21] P. Ferragina and A. Gulli, "A personalized search engine based on Web-snippet hierarchical clustering," in *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*, pp. 801–810, Chiba, Japan, May 2005.
- [22] D. L. Chan, R. W. P. Luk, W. K. Mak, H. V. Leong, E. K. S. Ho, and Q. Lu, "Multiple related document summary and navigation using concept hierarchies for mobile clients," in *Proceedings of the ACM Symposium on Applied Computing (SAC '02)*, pp. 627–633, Madrid, Spain, March 2002.
- [23] C. Carpineto, A. Della Pietra, S. Mizzaro, and G. Romano, "Mobile clustering engine," in *Proceedings of the Annual European Conference on Information Retrieval (ECIR '06)*, pp. 155–166, London, UK, April 2006.
- [24] G. Buchanan, M. Jones, and G. Marsden, "Exploring small screen digital library access with the greenstone digital library," in *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL '02)*, vol. 2458 of *Lecture Notes in Computer Science*, pp. 583–596, Rome, Italy, September 2002.
- [25] A. K. Karlson, G. Robertson, D. C. Robbins, M. Czerwinski, and G. Smith, "FaThumb: a facet-based interface for mobile search," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, vol. 2, pp. 711–720, Montreal, Canada, April 2006.
- [26] M. Käki, *Enhancing Web search result access with automatic categorization*, Ph.D. dissertation, Department of Computer Sciences, University of Tampere, Tampere, Finland, 2005.
- [27] M. Jones, G. Buchanan, and H. Thimbleby, "Sorting out searching on small screen devices," in *Proceedings of 4th International Symposium on Human Computer Interaction with Mobile Devices (MobileHCI '02)*, vol. 2411 of *Lecture Notes in Computer Science*, pp. 555–567, Pisa, Italy, September 2002.
- [28] M. Jones and G. Marsden, *Mobile Interaction Design*, John Wiley & Sons, Glasgow, UK, 2005.
- [29] E. W. De Luca and A. Nürnberger, "Supporting information retrieval on mobile devices," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '05)*, pp. 347–348, Salzburg, Austria, September 2005.
- [30] T. Paek, S. Dumais, and R. Logan, "WaveLens: a new view onto Internet search results," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '04)*, pp. 727–734, Vienna, Austria, April 2004.
- [31] S. Jones, M. Jones, and S. Deo, "Using keyphrases as search result surrogates on small screen devices," *Personal and Ubiquitous Computing*, vol. 8, no. 1, pp. 55–68, 2004.
- [32] S. O. Sweeney, F. Crestani, and A. Tombros, "Mobile delivery of news using hierarchical query-biased summaries," in *Proceedings of the ACM Symposium on Applied Computing (SAC '02)*, pp. 634–639, Madrid, Spain, March 2002.
- [33] R. W. White, I. Ruthven, and J. M. Jose, "Finding relevant documents using top ranking sentences: an evaluation of two alternative schemes," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*, pp. 57–64, Tampere, Finland, August 2002.
- [34] T. Heimonen and N. Jhaveri, "Visualizing query occurrence in search result lists," in *Proceedings of the 9th International Conference on Information Visualisation (IV '05)*, pp. 877–882, London, UK, July 2005.
- [35] K. Rodden, N. Milic-Frayling, R. Sommerer, and A. Blackwell, "Effective Web searching on mobile devices," in *Proceedings of the 17th Annual Conference on Human-Computer Interaction (HCI '03)*, pp. 281–296, Bath, UK, September 2003.

- [36] D. E. Rose and D. Levinson, “Understanding user goals in Web search,” in *Proceedings of the 13th International World Wide Web Conference (WWW '04)*, pp. 13–19, New York, NY, USA, May 2004.
- [37] Nokia 6680: Device Details—Nokia 6680, <http://forum.nokia.com/devices/6680>.
- [38] M. Käki, “Proportional search interface usability measures,” in *Proceedings of the 3rd Nordic Conference on Human-Computer Interaction*, pp. 365–372, Tampere, Finland, October 2004.
- [39] K. Hornbæk and E. L.-C. Law, “Meta-analysis of correlations among usability measures,” in *Proceedings of Conference on Human Factors in Computing Systems (CHI '07)*, pp. 617–626, San Jose, Calif, USA, April-May 2007.
- [40] T. Heimonen, A. Aula, H. Hutchinson, and L. Granka, “Comparing the user experience of search user interface designs,” in *Proceedings of Workshop on User Experience Evaluation Methods in Product Development (CHI '08)*, Florence, Italy, April 2008.